BioMol Concepts, Vol. 2 (2011), pp. 293–303 • Copyright © by Walter de Gruyter • Berlin • Boston. DOI 10.1515/BMC.2011.023

Review

DNA binding proteins: outline of functional classification

Zhiming Zheng[1,2] and Ya Wang[1,*]

[1] Department of Radiation Oncology, Emory University School of Medicine, Winship Cancer Institute of Emory University, Atlanta, GA 30322, USA
[2] Department of Neurosurgery, Provincial Hospital affiliated with Shandong University, Shandong University, 324, Jingwu Road, Jinan 250021, P.R. China

*Corresponding author
e-mail: ywang94@emory.edu

Abstract

DNA-binding proteins composed of DNA-binding domains directly affect genomic functions, mainly by performing transcription, DNA replication or DNA repair. Here, we briefly describe the DNA-binding proteins according to these three major functions. Transcription factors that usually bind to specific sequences of DNA could be classified based on their sequence similarity and the structure of the DNA-binding domains, such as basic, zinc-coordinating, helix-turn-helix domains, etc. Most DNA replication factors do not need a specific sequence of DNA, but instead mainly depend on a DNA structure, with the exception of the origin recognition complex in yeast or *Escherichia coli* that recognizes the DNA sequences at particular origins. DNA replication includes initiation and elongation. The major DNA-binding proteins involved in these two steps are briefly described. DNA repair proteins bound to DNA depend on the damaged DNA structure. They are classified to base excision repair, DNA mismatch repair, nucleotide excision repair, homologous recombination repair and non-homologous end joining. The major DNA-binding proteins involved in these pathways are briefly described. Histone and high mobility group are two examples of DNA-binding proteins that do not belong to the three categories above and are briefly described. Finally, we warn that the non-specific binding proteins might have an affinity to some non-specific medium materials such as protein A or G beads that are commonly used for immune precipitation, which can easily generate false positive signals while detecting protein-protein interaction; therefore, the results need to be carefully analyzed using positive/negative controls.

Keywords: DNA binding protein; DNA repair; DNA replication; Ku; p53; transcription.

Introduction

DNA-binding proteins directly regulate genomic functions, such as transcription, replication and repair because all genes are encoded in DNA. These proteins are composed of DNA-binding domains and have a specific or general affinity to recognize and bind to either single- or double-stranded DNA. DNA-binding proteins have common folding patterns known as DNA-binding motifs. Each DNA-binding motif is composed of a recognition region and a stabilization region. Recognition of DNA by protein can take place at two levels: (i) non-specific binding – between protein side chains and a DNA sugar/phosphate backbone and (ii) specific binding – between protein side chains and nucleotide bases. The physiological role of DNA-binding proteins is determined by the affinity and specificity of the DNA-protein interaction. These properties depend upon precise interactions between amino acids in the protein and the structure or nucleotides in the DNA site. DNA-binding proteins are among the most widespread cellular proteins; therefore, it is impossible to describe the whole DNA-binding protein picture in one review. Previously, there were some review papers that described DNA-binding proteins according to the specific domains of the protein (1–4) or according to the bound DNA structures of the protein (5, 6). Here, we attempt to briefly describe DNA-binding proteins in three major categories: transcription, replication and repair, even though many DNA-binding proteins have overlapping functions. For example, replication protein A (RPA) is an important DNA replication factor and also plays an important role in DNA repair. Some important DNA-binding proteins such as histone and high mobility group (HMG) that are not easy to classify into these three categories are outlined at the end of this review. RNA-binding proteins, such as RNA polymerases, termination factor (rho) and anti-termination factors (lambda phage, etc.), as well as DNA modification enzymes, such as methylase and hydroxylase are not included in this review. In addition, we briefly classified the DNA-binding proteins into these three categories but did not intend to describe the whole process of transcription, replication and repair. The purpose of this review is to help readers quickly find an interesting DNA-binding protein in a major function field. We tried to cover the most important DNA-binding proteins; however, owing to our limited knowledge and manuscript space, we might have still missed some important DNA-binding proteins.

DNA-binding proteins: transcription factors

These classical proteins are the largest group among DNA-binding proteins. These transcription factors bind to a specific sequence and control the rate of gene transcription. They regulate the transcription rate alone or with other proteins in a complex, by promoting (as an activator) or blocking

(as a repressor) the recruitment of RNA polymerase to specific genes (7, 8). In addition to the regulation coding gene expression, transcription factors also regulate non-coding RNA, such as rRNA, tRNA, microRNA and other long non-coding RNA expression in a similar mechanism. The detailed transcription process has been described in other reviews (7, 8). Here, we list the transcription factors in several groups (Tables 1–4) based on the sequence similarity and, hence, the tertiary structure of the DNA-binding domains (9, 10). p53 as an important human transcription factor is described in detail for its DNA-binding property. Table 5 shows the examples of some transcription factors that recognize specific DNA sequences in the different groups. Table 6 shows some prokaryotic transcription factors that are not included in Tables 1–4.

p53 has been described as the 'guardian of the genome', referring to its role in conserving stability by preventing genome mutation (12). p53 in humans is encoded by the *TP53* gene located on the short arm of chromosome 17 (17p13.1) (13–16). p53 as a tumor suppressor regulates the cell cycle, apoptosis and responds to DNA damage. There are some outstanding reviews describing the tumor suppressor functions of p53 (17, 18). Here, we only address the most important transcription features of p53 according to its DNA-binding property. p53 as a transcription factor recognizes the consensus sequence that is 5′-RRRCWWGYYY-N (0–13)-RRRCWWGYYY-3′ (Figure 1). This region is located in the regulatory regions of the target that is activated by p53. The presence of p53 response elements in or around genes (promoters, upstream sequences, introns) is a powerful predictor of regulation and activation of a particular gene by p53 (19). For example, the $p21^{WAF1/Cip1}$ gene contains several p53 response elements that mediate direct binding of the p53 protein, resulting in transcriptional activation of the gene encoding the $p21^{WAF1,\ Cip1}$ protein following DNA damage (20). The $p21^{WAF1/Cip1}$ protein binds directly to cyclin-CDK complexes and inhibits their kinase activity, thereby, causing a cell cycle arrest that allows repair to take place (21). Human p53 contains 393 amino acids and has seven domains: 1) an acidic N-terminus transcription-activation domain (TAD), also known as activation domain (AD) 1, which activates transcription factors at residues 1–42. The N-terminus contains two complementary transcriptional activation domains, with a major one at residues 1–42 and a minor one at residues 55–75, which specifically involves the regulation of several pro-apoptotic genes (22), 2) AD 2 is important for

Table 1 Transcription factors with basic domains.

Basic domains (basic-helix-loop-helix)	Transcription factor (specific functions)
Basic helix-loop-helix (bHLH)	MyoD (myogenic transcription factors)
	Achaete-Scute: a group of four genes: achaete, scute, lethal of scute and asense in the fruit fly (regulation of nervous system development)
	Tal/Twist/Atonal/Hen family: lymphoid factors, mesodermal Twist-like factors, HEN, Mesp, atonal, pancreatic factors, etc.
	Hairy family: Hairy/E(SPL), fungal regulators
Basic-leucine zipper (bZIP)	AP-1(-like) components including c-Fos/c-Jun, Maf, NF-E2, fungal regulators, CRE-BP/ATF, etc.
	CREB family
	C/EBP-like factor family
	Plant G-box binding factors, etc.
	bZIP/PAR
	AREB/ABF
	ZIP only
Basic helix-loop-helix/basic-leucine zipper (bHLH-ZIP)	Ubiquitous bHLH-ZIP factors including USF1, USF2, SREBP, c-Myc, Mad/Max, etc.
Other basic domain	NF1 including A, B, C, X, etc.
	RF-X including 1, 2, 3, 4, 5 and ANK, etc.
	AP-2 (bHSH)

Table 2 Transcription factors with zinc-coordinating domains.

Zinc-coordinating domains	Transcription factors
Cys4 zinc finger of nuclear receptor type	Steroid hormone receptor-like, thyroid hormone receptor-like, hepatocyte nuclear factor 4-like, estrogen-like, nerve growth factor IB-like, Fushi tarazu-F1-like, germ cell nuclear factor, Knirps/DAX 1-like, Trithorax, etc.
Diverse Cys4 zinc finger	GATA factors
Cys2His2 zinc finger	Ubiquitous factors including TFIIIA, Sp1, etc.; developmental/cell cycle regulators including Egr/Krox, Kruppel-like, GLI-like, etc.; metabolic regulators in fungi, large factors with NF-6B-like binding properties
Cys6	Metabolic regulators in fungi
Cys4HisCys3	PHD fingers

Table 3 Transcription factors with helix-turn-helix domains.

Helix-turn-helix domains	Transcription factors (specific functions)
Homeodomain only	AbdB, Antp, Cad, Cut, Dll, Ems, En, Eve, PBC, Prd, HD-ZIP, H2.0, HNF1, Msh, NK-2, Ubx
POU domain factors	Oct, etc.
Homeodomain with LIM region (cysteine-rich motif)	Chip, etc.
Homeodomain plus zinc finger motifs	ZEB-1, Zfh-1, etc.
Paired plus homeodomain	PAX3-FKHR, etc.
Fork head/winged helix	E2F, DP, Trident, etc.
Tryptophan cluster	Myb, Ets-type, interferon regulator factors, etc.
TEA (transcriptional enhancer factor)	TEAD1, 2, 3, 4, etc.
Heat shock factors (HSFs)	HSF-1, etc.
Prokaryotic type helix-turn-helix domains	
AraC family	Rv1931c, etc.
LysR family	NtcB, etc.
DeoR family	lacR1, lacR2, fruR, etc.
LacI family	ccpA (catabolite control protein A), malR (lacI family transcription regulator), ribose operon repressor, scrR (sucrose operon repressor)
TetR/AcrR family	PigZ, etc.
Lrp/AsnC family	Rv2779c, etc.
LuxR family	TraR, etc.
MarR family	MarR family transcriptional regulator, etc.
Fur family	perR (ferric uptake regulation protein)
IclR family	MhpR, etc.
CRP/FNR family	PrfA, etc.
Rrf2 family	Slr0846, etc.
Other families	scar (iron-dependent repressor), PadR family transcriptional regulator, birA (biotin-protein ligase), codY (transcriptional repressor), Cro/CI family transcriptional regulator

Table 4 Transcription β-scaffold factors with minor groove contact domains.

β-Scaffold factors with minor groove contact domains	Transcription factors
Rel homology region (RHR)	Rel/ankyrin, NF-κB, ankyrin, nuclear factor of activated T-cells (NFATC) including NFATC1, 2, 3, etc.
STAT family	STAT factors
MADS box	MEF2, homeotic genes, yeast regulators, SRF (serum response factor), etc.
p53 family	p53 (Figure 1), p63, p73, etc.
TATA binding (11)	TBP, SOX genes (SRY), TCF-1, HMG2-related (SSRP1), MATA, UBF, other HMG box factors
Cold-shock domain	DbpA-like, YB-1/DbpB-like, FRG Y2-like
Heteromeric CCAAT factors	CBF-A, CBF-B, etc.
Grainyhead	Grainyhead-like factors
Runt	PEBP2/PEA2

TBP, TATA-binding protein.

Table 5 Examples of some transcription factors and their recognized DNA sequences.

Factor	Structural type	Recognition sequence	Binds as
SP1	Zinc finger	5′-GGGCGG-3′	Monomer
AP-1	Basic zipper	5′-TGA(G/C)TCA-3′	Dimer
C/EBP	Basic zipper	5′-ATTGCGCAAT-3′	Dimer
Heat shock factor	Basic zipper	5′-XGAAX-3′	Trimer
ATF/CREB	Basic zipper	5′-TGACGTCA-3′	Dimer
c-Myc	Basic-helix-loop-helix	5′-CACGTG-3′	Dimer
Oct-1	Helix-turn-helix	5′-ATGCAAAT-3′	Monomer
NF-1	Novel	5′-TTGGCXXXXXGCCAA-3′	Dimer

(G/C)=G or C, X=A, T, G or C.

Table 6 Other prokaryotic transcription factors.

BglG family	licT (antiterminator), PTS system, mannitol (cryptic)-specific, etc.
Others	argR1, argR2 (arginine repressor), ahrC2 (arginine repressor), hrcA (heat-inducible transcription repressor), ctsR (transcriptional regulator), ParB (chromosome partitioning protein), PyrR (bifunctional pyrimidine regulatory protein PyrR uracilphosphoribosyl transferase), NrdR (transcriptional regulator), PurR (pur operon repressor), ComX2 (competence-specific sigma factor)

apoptotic activity at residues 43–63, 3) the proline-rich domain is important for the apoptotic activity of p53 at residues 64–92, 4) the central DNA-binding core domain (DBD) (Figure 1) contains one zinc atom and several arginine amino acids at residues 100–300, 5) the nuclear localization signaling domain contains residues 316–325, 6) the homo-oligomerization domain contains residues 307–355. Tetramerization is essential for the activity of p53 *in vivo*, 7) C-terminal involves down-regulation of DNA binding of the central domain at residues 356–393 (23). p53 mutation in the DBD that deactivate p53 usually occur in cancer. Most of these mutations destroy the ability of the protein to bind to its target DNA sequences, and thus prevents transcriptional activation of these genes. As such, mutations in the DBD are recessive loss-of-function mutations. p53 mutation in the oligomerization domain (OD) dimerizing with wild-type p53 prevent p53 from activating transcription. Therefore, OD mutations have a dominant-negative effect on the function of p53. p53 has two family members: p63 and p73. These p53 family proteins have overlapping and distinct functions (24).

Other transcription factors

In addition to most transcriptional factors that could be grouped as described in Tables 1–4, there are some transcriptional factors that are not easy to classify into any group. The following transcriptional factors are such examples: Copper fist proteins; HNGI(Y) family; Pocket domain transcription factors; E1A-like factors; AP2/EREBP-related factors, EREBP, AP2/B3, ARF family, ABI family, RAV family, etc. There are other prokaryotic transcription factors that are summarized in Table 6.

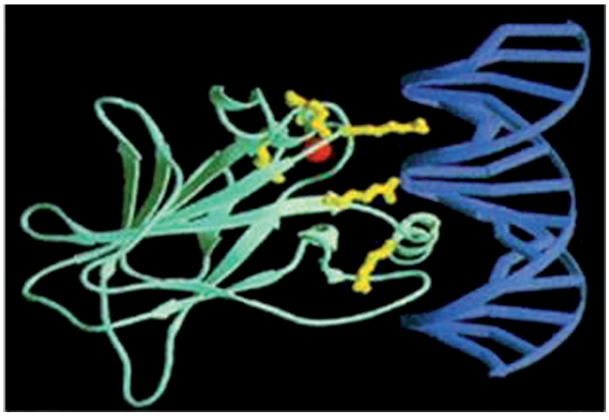

Figure 1 The structure of the core domain of the p53 protein is shown in light blue bound to DNA (dark blue).
The six most frequently mutated amino acids in human cancer are shown in yellow – all are residues important for p53 binding to DNA. Red ball: zinc atom. Reproduced with permission from Ref. (94).

DNA-binding protein: DNA replication factors

DNA replication is the essential requirement for a cell to divide (25). This process is completed by a group of proteins working on the DNA template. Different from the transcription factors, most DNA replication factors do not need a specific sequence of DNA but instead mainly depend on the DNA structure, except the origin recognition complex (ORC) (in yeast) and dnaA (in *Escherichia coli*) that recognize the DNA sequences at particular points, known as 'origins' (26). These initiator proteins recruit other proteins to separate the two strands, forming a bubble and initiating replication forks (27). Origins tend to be 'AT-rich' (rich in adenine and thymine bases) that assist in this process, because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair) (28). Once strands are separated, RNA primers are created on the template strands. To be more specific, the leading strand receives one RNA primer per active origin of replication, whereas the lagging strand receives several; these several fragments of RNA primers found on the lagging strand of DNA are called Okazaki fragments (named after their discoverer). The single-strand DNA-binding proteins (SSB in *E. coli* and RPA in eukaryotic cells) are required for DNA replication (5, 29). DNA polymerase extends the leading strand in one continuous motion and the lagging strand in a discontinuous motion. RNase removes the RNA fragments used to initiate replication by DNA polymerase, and another DNA polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill in these nicks, thus completing the newly replicated DNA molecule. As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming two replication forks. In bacteria, which have a single origin of replication on their circular chromosome, this process eventually creates a 'theta structure' (resembling the Greek letter theta: θ). By contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these (30).

After DNA replication forks form, one of the two DNA strands is named as the leading strand, and the replication fork moves along it in a 3′-to-5′ direction. The other one of the two DNA strands is named as the lagging strand, and the replication fork moves along it in a 5′-to-3′ manner. The

leading strand allows the new strand to be complementary synthesized 5′-to-3′ in the same direction as the movement of the replication fork. On the leading strand, a polymerase 'reads' the DNA and continuously adds nucleotides to it. This polymerase is DNA polymerase III (DNA pol III) in prokaryotes and presumably pol ε (31, 32) in eukaryotes. In contrast to the leading strand, because of its orientation, which is opposite to the working orientation of DNA pol III and moves on a template in a 3′-to-5′ manner, replication of the lagging strand is more complicated than that of the leading strand. On the lagging strand, primase 'reads' the DNA and adds RNA to it in short, separated segments. In eukaryotes, primase is intrinsic to pol α (33). DNA pol III or pol δ lengthens the primed segments forming Okazaki fragments. Primer removal in eukaryotes is also performed by pol δ. In prokaryotes, DNA pol I 'reads' the fragments, removes the RNA using its flap endonuclease domain (RNA primers are removed by 5′-3′ exonuclease activity of polymerase I) (34) and replaces the RNA nucleotides with DNA nucleotides (this is necessary because RNA and DNA use slightly different types of nucleotides). DNA ligase joins the fragments together.

The dynamics at the replication fork requires helicase to unwind the DNA at the site and the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead (35). This build-up would form a resistance that would eventually halt the progress of the replication fork. DNA topoisomerases are enzymes that solve these physical problems in the coiling of DNA. Topoisomerase I cuts a single backbone on the DNA, enabling the strands to swivel around each other to remove the build-up of twists. Topoisomerase II cuts both backbones, enabling one double-stranded DNA to pass through another; thereby, removing knots and entanglements that can form within and between DNA molecules (36). Bare single-stranded DNA has a tendency to fold back into itself and form secondary structures; these structures can interfere with the movement of DNA polymerase. To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation (25). Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase to maintain contact with its template; thereby, assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between the template and RNA primers (37).

The DNA-binding proteins in DNA replication can be mainly classified according to the two major processes of DNA replication.

DNA replication initiation

In prokaryotic cells, pre-replication complex (pre-RC) forms at the origin of replication during the initiation step of DNA replication. The proteins involved in the pre-RC are essential for DNA replication. In prokaryotes, the pre-RC is made up of the following factors: a replication initiation factor such as dnaA, a primase such as dnaG (generates a RNA primer to be used in DNA replication) and a DNA holoenzyme (actually a complex of enzymes that performs the actual replication). In eukaryotes, the pre-RC is made up of the following factors: a six-subunit complex (ORC binds to the origin), two regulatory proteins called Cdc6 and Cdt1 (recruited by ORC), the mini-chromosome maintenance proteins (MCMs) and the putative helicase complex. These proteins assemble on cellular origins in the G1 phase of the cell cycle. Once these proteins are assembled, the MCMs are phosphorylated and DNA replication begins.

DNA replication elongation

In prokaryotic cells, the elongation requires DNA pol III holoenzyme (dnaC, dnaE, dnaH, dnaN, dnaQ, dnaT, dnaX), Replisome, DNA ligase, DNA clamp, Topoisomerase (DNA gyrase), DNA pol I (Klenow fragment) and DNA ligase. In eukaryotic cells, the synthesis occurs in the S-phase and requires replication factor C (RFC1), flap endonuclease (FEN1), topoisomerase, RPA, eukaryotic DNA polymerase: delta (POLD1, POLD2, POLD3, POLD4), DNA clamp (PNCA) and DNA ligase.

DNA-binding protein: DNA repair factors

Many DNA repair processes involved in removing the damaged DNA and filling the gap share some factors with DNA replication. DNA repair factors are generally classified into different groups according to repairing different types of DNA damage.

Base excision repair (BER)

BER is primarily responsible for removing small, non-helix-distorting base lesions from the genome. The most common types of base damage include deamination, oxidation and alkylation. These modifications can affect the ability of the base to hydrogen-bond, resulting in incorrect base pairing and, as a consequence, mutations in the DNA. BER is initiated by DNA glycosylases that include Ogg1 (recognizes 8-oxoguanine), Mag1 (recognizes 3-methyladenine) and UNG (removes uracil from DNA). DNA glycosylase recognizes and removes specific damaged or inappropriate bases, forming abasic (AP) sites. These are then cleaved by an AP endonuclease to yield a 3′ hydroxyl adjacent to a 5′ deoxyribosephosphate (dRP). In humans, only a single AP endonuclease, APE1 (a member of the ExoIII family), has been identified (38). The resulting single-strand break can then be processed by either short-patch (where a single nucleotide is replaced) or long-patch BER (where 2–10 new nucleotides are synthesized) (39). For ligation to occur, a DNA strand break must have a hydroxyl on its 3′ end and a phosphate on its 5′ end. In humans, polynucleotide kinase-phosphatase (PNKP) promotes formation of these ends during BER. The AP endonuclease also participates in the 3′ end processing. They possess 3′ phosphodiesterase activity

for removing a variety of 3′ lesions including phosphates, phosphoglycolates and aldehydes. DNA polymerase β (pol β) is the main human polymerase that catalyzes short-patch BER, with pol λ able to compensate in its absence (40). These polymerases are members of the pol X family and typically insert only a single nucleotide. In addition to polymerase activity, these enzymes have a lyase domain that removes the 5′ dRP left behind by AP endonuclease cleavage. During long-patch BER, DNA synthesis is thought to be mediated by pol δ and pol ε along with the processivity factor proliferating cell nuclear antigen (PCNA), the same polymerases that carry out DNA replication. These polymerases perform displacing synthesis, meaning that the downstream 5′ DNA end is 'displaced' to form a flap. FEN1 removes the 5′ flap generated during long-patch BER. This endonuclease shows a strong preference for a long 5′ flap adjacent to a 1-nt 3′ flap (41). In addition to its role in long-patch BER, FEN1 cleaves flaps with a similar structure during Okazaki fragment processing, an important step in lagging strand DNA replication. DNA ligase III along with its cofactor XRCC1 catalyzes the nick-sealing step in short-patch BER in humans. DNA ligase I ligates the break in long-patch BER.

DNA mismatch repair

This is a system for recognizing and repairing erroneous insertion, deletion and misincorporation of bases that can arise during DNA replication and recombination, as well as repairing some forms of DNA damage (42, 43). Mismatch repair is strand-specific. The mismatch repair machinery distinguishes the newly synthesized strand from the template (parental). In Gram-negative bacteria, transient hemimethylation distinguishes the strands (the parental is methylated and the daughter is not). It is suspected that in eukaryotes, newly synthesized lagging-strand DNA transiently contains nicks (before being sealed by DNA ligase) and provides a signal that directs mismatch proofreading systems to the appropriate strand. This implies that these nicks must be present in the leading strand, but it is unclear how. The mismatch repair is a highly conserved process from prokaryotes to eukaryotes. The gene products are therefore called the 'Mut' (denotes mutator) proteins and are the major active components of the mismatch repair system. Three of these proteins are essential in detecting the mismatch and directing repair machinery to it: MutS, MutH and MutL (MutS is a homolog of HexA and MutL of HexB). MutS forms a dimer ($MutS_2$) that recognizes the mismatched base on the daughter strand and binds the mutated DNA. MutH binds at hemimethylated sites along the daughter DNA, but its action is latent, being activated only upon contact by a MutL dimer ($MutL_2$) which binds the MutS-DNA complex and acts as a mediator between $MutS_2$ and MutH, activating the latter. The DNA is looped out to search for the nearest d (GATC) methylation site to the mismatch, which could be up to 1 kb away. Upon activation by the MutS-DNA complex, MutH nicks the daughter strand near the hemimethylated site and recruits a UvrD helicase (DNA helicase II) to separate the two strands with a specific 3′-to-5′ polarity. The entire MutSHL complex then slides along the DNA in the direction of the mismatch, liberating the strand to be excised as it goes. An exonuclease trails the complex and digests the single-strand DNA tail. The exonuclease recruited is dependent on which side of the mismatch MutH incises the strand: 5′ or 3′. If the nick made by MutH is on the 5′ end of the mismatch, either RecJ or ExoVIII (both 5′-to-3′ exonucleases) is used. If, however, the nick is on the 3′ end of the mismatch, ExoI (a 3′-to-5′ enzyme) is used.

Nucleotide excision repair (NER)

DNA constantly requires repair owing to damage that can occur to bases from a vast variety of sources including chemicals, radiation, etc. (44). NER is a particularly important mechanism by which the cell can prevent unwanted mutations by removing the vast majority of UV-induced DNA damage (mostly in the form of thymine dimers and 6-4 photoproducts). The importance of this repair mechanism is evidenced by the severe human diseases that result from inborn genetic mutations of NER proteins including Xeroderma pigmentosum and Cockayne syndrome. Recognition of bulky distortions in the shape of the DNA double helix leads to the removal of a short single-stranded DNA segment that includes the lesion, creating a single-strand gap in the DNA, which is subsequently filled in by DNA polymerase, which uses the undamaged strand as a template. For Uvr proteins, the process of NER is controlled in *E. coli* by the UvrABC endonuclease enzyme complex, which consists of four Uvr proteins: UvrA, UvrB, UvrC and DNA helicase II (sometimes also known as UvrD in this complex). First, a UvrA-UvrB complex scans the DNA, with the UvrA subunit recognizing distortions in the helix caused, for example, by pyrimidine dimers. When the complex recognizes such a distortion, the UvrA subunit leaves and a UvrC protein comes in and binds to the UvrB monomer and, hence, forms a new UvrBC dimer. UvrB cleaves a phosphodiester bond, four nucleotides downstream of the DNA damage, and the UvrC cleaves a phosphodiester bond eight nucleotides upstream of the DNA damage and creates a 12 nucleotide excised segment. DNA helicase II (sometimes called UvrD) then comes in and removes the excised segment by actively breaking the hydrogen bonds between the complementary bases. The resultant gap is then filled in using DNA pol I and DNA ligase. The basic excision process is very similar in higher cells, but these cells usually involve many more proteins. There are nine major proteins involved in NER in mammalian cells and their names come from the diseases associated with the deficiencies in those proteins. XPA, XPB, XPC, XPD, XPE, XPF and XPG are derived from Xeroderma pigmentosum, and CSA and CSB (45) that represent proteins linked to Cockayne Syndrome. Additionally, the proteins ERCC1, RPA, RAD23A, RAD23B and others also participate in NER. The resulting gap in DNA is filled by DNA pol δ or ε by copying the undamaged strand. PCNA assists the DNA polymerase in the reaction, and RPA protects the other DNA strand from degradation during NER. Finally, DNA ligase seals the nicks to finish NER.

Homologous recombination repair (HRR)

HRR is a type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. It is most widely used by cells to accurately repair harmful breaks that occur on both strands of DNA, known as double-strand breaks. After a double-strand break (DSB) occurs, the Mre11-Rad50-Xrs2 (MRX) complex in yeast and the Mre11-Rad50-Nbs1 (MRN) complex in mammalian is thought to promote bridging of the DNA ends (46, 47). Next a resection, in which DNA around the 5′ ends of the break is cut back, is carried out in two distinct steps. In the first step of resection, the MRX complex recruits the Sae2 protein. The two proteins then trim back the 5′ ends on either side of the break to create short 3′ overhangs of single-strand DNA. In the second step, 5′→3′ resection is continued by the Sgs1 helicase and the Exo1 and Dna2 nucleases. As a helicase, Sgs1 'unzips' the double-strand DNA, whereas Exo1 and the nuclease activity of Dna2 allows them to cut the single-stranded DNA produced by Sgs1 (48). The RPA protein, which has high affinity for single-stranded DNA, then binds the 3′ overhangs (5). With the help of several other proteins that mediate the process, the Rad51 protein (and Dmc1, in meiosis) then forms a filament of nucleic acid and protein on the single strand of DNA coated with RPA. This nucleoprotein filament then begins searching for DNA sequences similar to that of the 3′ overhang. After finding such a sequence, the single-stranded nucleoprotein filament moves into (invades) the similar or identical recipient DNA duplex in a process called 'strand invasion'. In cells that divide through mitosis, the recipient DNA duplex is generally a sister chromatid, which is identical to the damaged DNA molecule and provides a template for repair. In meiosis, however, the recipient DNA tends to be from a similar but not necessarily identical homologous chromosome (49). A displacement loop (D-loop) is formed during strand invasion between the invading 3′ overhang strand and the homologous chromosome. After strand invasion, a DNA polymerase extends the end of the invading 3′ strand by synthesizing new DNA. This changes the D-loop to a cross-shaped structure known as a Holliday junction. Following this, more DNA synthesis occurs on the invading strand (i.e., one of the original 3′ overhangs), effectively restoring the strand on the homologous chromosome that was displaced during strand invasion (49). After the strands anneal, a small flap of DNA can sometimes remain. Any such flaps are removed, and the SDSA pathway finishes with the resealing, also known as 'ligation', of any remaining single-stranded gaps (50).

Non-homologous end joining (NHEJ)

NHEJ is a pathway that repairs DSBs in DNA. NHEJ is referred to as 'non-homologous' because the break ends are directly ligated without the need for a homologous template. The term 'non-homologous end joining' was coined in 1994 by Pfeiffer and colleagues (51). NHEJ typically utilizes short homologous DNA sequences called microhomologies to guide repair. These microhomologies are often present in single-stranded overhangs on the ends of DSBs. When the overhangs are perfectly compatible, NHEJ usually repairs the break accurately (52–55). Imprecise repair leading to loss of nucleotides can also occur, but is much more common when the overhangs are not compatible. Inappropriate NHEJ can lead to translocations and telomere fusion, hallmarks of tumor cells (56). NHEJ is evolutionarily conserved throughout all kingdoms of life and is the predominant DSB repair pathway in mammalian cells (57). In yeast, however, homologous recombination dominates when the organism is grown under common laboratory conditions and many species of bacteria lack an end joining pathway and thus rely completely on homologous recombination to repair DSBs. NHEJ in eukaryotes utilizes several proteins, that participate in the following steps: the Ku heterodimer is an abundant and multifunction DNA-binding protein containing two subunits, Ku80 and Ku70, with ATPase and possible helicase activity. Both subunits have a strong ability to bind to double-strand DNA ends (Figure 2) and promote DNA DSBs, NHEJ repair and the process of variable (diversity) joining recombination in mammalian cells, which contributes to genomic integrity through its ability to bind DNA DSBs and facilitate repair by the NHEJ pathway. The Ku heterodimer forms a complex with DNA-PKcs that is present in mammals but absent in yeast. Ku is a basket-shaped molecule that slides onto the DNA end and translocates inward (58). Ku can function as a docking site for other NHEJ proteins and is known to interact with the DNA ligase IV complex and XLF (59, 60). Little is known about the function of nucleases in NHEJ. Artemis is required for opening the hairpins that are formed on DNA ends during variable, diversity, joining [V(D)J] recombination, a specific type of NHEJ, and can also participate in end trimming during general NHEJ (61). The X family DNA polymerases pol λ and pol μ (pol4 in yeast) fill gaps during NHEJ (54, 62, 63). Yeast lacking pol4 are unable to join 3′ overhangs that require gap filling, but remain proficient for gap filling at 5′ overhangs (63). This is because the primer

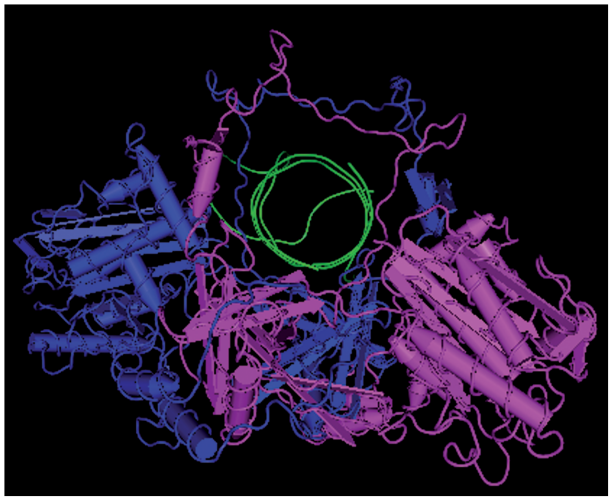

Figure 2 Crystal structure of human Ku bound to DNA. Ku70 is shown in purple, Ku80 in blue and the DNA strand in green. Reproduced with permission from Ref. (58).

terminus used to initiate DNA synthesis is less stable at 3′ overhangs, necessitating a specialized NHEJ polymerase. The DNA ligase IV complex, consisting of the catalytic subunit DNA ligase IV and its cofactor XRCC4 (Dnl4 and Lif1 in yeast), performs the ligation step of repair (64). XLF, also known as Cernunnos, is homologous to yeast Nej1 and is also required for NHEJ (65, 66). Although the precise role of XLF is unknown, it interacts with the XRCC4/DNA ligase IV complex and probably participates in the ligation step (67). Recent evidence suggests that XLF re-adenylates DNA ligase IV after ligation, recharging the ligase and allowing it to catalyze a second ligation (68). In addition to the DNA-PK dependent NHEJ, the alternated pathways of NHEJ (69, 70) have been described with important functions in B cell development and implicated in DSB repair and cancer formation (71) repair of DNA DSBs, NHEJ is also essential for the V(D)J recombination (72–74) and the maintenance of telomeres (75).

Other important DNA-binding proteins

There are some important DNA-binding proteins that are difficult to classify into the categories of transcription, replication or repair but are important for cell function. Here, we briefly describe two of them: histone and HMG.

Histones are highly alkaline proteins found in eukaryotic cell nuclei, which package and order the DNA into structural units called nucleosomes (76, 77). Histones ‘are highly conserved and can be grouped into five major classes: H1/H5, H2A, H2B, H3 and H4’ (77–79). These are organized into two superclasses as core histones (H2A, H2B, H3 and H4) and linker histones (H1 and H5). The human histones are classified as described in Table 7. Two of each of the core histones assemble to form one octameric nucleosome core particle by wrapping 147 base pairs of DNA around the protein spool in a 1.65 left-handed superhelical turn (80). The linker histone H1 binds the nucleosome and the entry and exit sites of the DNA, thus locking the DNA into place (81) and allowing the formation of higher order structure. This involves the wrapping of DNA around nucleosomes with approximately 50 base pairs of DNA separating each pair of nucleosomes (also referred to as linker DNA). The assembled histones and DNA is called chromatin. Higher order structures include the 30 nm fiber (forming an irregular zigzag) and 100 nm fiber, these being the structures found in normal cells. During mitosis and meiosis, the condensed chromosomes are assembled through interactions between nucleosomes and other regulatory proteins. The nucleosome core is formed into two H2A-H2B dimers and a H3-H4 tetramer, forming two nearly symmetrical halves by tertiary structure (C2 symmetry; one macromolecule is the mirror image of the other) (80). The H2A-H2B dimers and H3-H4 tetramer also show pseudodyad symmetry. The four ‘core’ histones (H2A, H2B, H3 and H4) are relatively similar in structure and are highly conserved through evolution, all featuring a ‘helix-turn-helix-turn-helix’ motif (which allows for easy dimerization). They also share the feature of long ‘tails’ on one end of the amino acid structure – this being the location of post-translational modification. In general, genes that are active have less bound histone, whereas inactive genes are highly associated with histones during interphase (82–84). It also appears that the structure of histones has been evolutionarily conserved, as any deleterious mutations would be severely maladaptive.

High mobility group (HMG) proteins

HMG proteins were arbitrarily classed as a specific type of non-histone protein associated with isolated chromatin. HMGs are ubiquitous nuclear proteins that regulate and facilitate various DNA-related activities such as transcription,

Table 7 Human histone classification.

Super family	Family	Subfamily	Members
Linker	H1	H1F	H1F0, H1FNT, H1FOO, H1FX
		H1H1	HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T
Core	H2A	H2AF	H2AFB1, H2AFB2, H2AFB3, H2AFJ, H2AFV, H2AFX, H2AFY, H2AFY2, H2AFZ
		H2A1	HIST1H2AA, HIST1H2AB, HIST1H2AC, HIST1H2AD, HIST1H2AE, HIST1H2AG, HIST1H2AI, HIST1H2AJ, HIST1H2AK, HIST1H2AL, HIST1H2AM
		H2A2	HIST2H2AA3, HIST2H2AC
	H2B	H2BF	H2BFM, H2BFO, H2BFS, H2BFWT
		H2B1	HIST1H2BA, HIST1H2BB, HIST1H2BC, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, HIST1H2BI, HIST1H2BJ, HIST1H2BK, HIST1H2BL, HIST1H2BM, HIST1H2BN, HIST1H2BO
		H2B2	HIST2H2BE
	H3	H3A1	HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H3G, HIST1H3H, HIST1H3I, HIST1H3J
		H3A2	HIST2H3C
		H3A3	HIST3H3
	H4	H41	HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G, HIST1H4H, HIST1H4I, HIST1H4J, HIST1H4K, HIST1H4L
		H44	HIST4H4

replication, recombination and repair (85). They bind to DNA and chromatin and act as 'architectural elements' that induce both short- and long-range changes in the structure of their binding sites. The functional motifs of the ubiquitous HMG proteins are widespread and found in the DNA-binding domains of numerous regulatory proteins. HMGs were subdivided into three groups named HMG-1/-2, HMG-14/-17 and HMG-I/Y. These proteins are considered as canonical HMG proteins. Subsequent studies revealed that the functional motifs characteristic of each of the canonical HMG proteins are widespread among nuclear proteins. Proteins containing any of the functional motifs of the canonical HMG proteins are called HMG-motif proteins. In fact, the canonical HMG proteins can be considered to be a subclass of the HMG-motif proteins (86). The HMG motif proteins are now subdivided into three superfamilies which are now named: 1) HMGB (previous name HMG-1/-2), 2) HMGN (previous name HMG-14/-17) and 3) HMGA (previous name HMG-I/Y/C). Each HMG family has a characteristic functional sequence motif. The functional motif of the HMGB family is named 'HMG-box', that of the HMGN family is named 'nucleosomal binding domain' and that of the HMGA family is named 'AT-hook'. Proteins containing any of these functional motifs embedded in their sequence are known as 'HMG-motif proteins'. HMGs affect the activities of various regulatory molecules including hormone receptors (87), p53 (88), the RAG proteins involved in V(D)J recombination (89), the homeotic protein HOXD9 (90), immunity (91) and tumors (92).

In addition to the above classification of the DNA-binding proteins, we would like to share our experiences: the non-specific DNA-binding proteins have the affinity to some non-specific medium materials such as protein A or protein G beads that are commonly used for immune precipitation (93). The properties of these DNA-binding proteins easily result in false-positive signals for protein-protein interaction and, therefore, need careful analysis using positive and negative controls.

In summary, in this review we briefly described DNA-binding proteins according to their functions: transcription, DNA replication, DNA repair (including BER, mismatch repair, NER, HRR and NHEJ) and other DNA-binding proteins that are difficult to classify into these three categories. Transcription factors are specific DNA-binding proteins that recognize specific DNA sequences. Most DNA replication factors do not need a specific sequence of DNA but instead mainly depend on the DNA structure, except the ORC (in yeast) and dnaA (in *E. coli*) that recognize the DNA sequences at particular points, known as origins. DNA replication could be divided into two major steps: initiation and elongation. DNA repair protein that binds to DNA do not depend on a DNA sequence but depend on the damaged DNA structure. There are different DNA repair pathways to repair different types of DNA damage: BER, DNA mismatch repair, NER, HRR and NHEJ, etc. Histones and HMGs are two examples of DNA-binding proteins that do not belong to the three categories but have important functions affecting multiple cell functions.

Acknowledgments

This work is supported by NIH grant (GM080771) to Y.W. We thank Ms. Doreen Theune for editing the manuscript.

References

1. Harrison SC. A structural taxonomy of DNA-binding domains. Nature 1991; 353: 715–9.
2. Wolffe AP, Tafuri S, Ranjan M, Familari M. The Y-box factors: a family of nucleic acid binding proteins conserved from Escherichia coli to man. New Biol 1992; 4: 290–8.
3. Luscombe NM, Thornton JM. Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J Mol Biol 2002; 320: 991–1009.
4. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol 2010; 48: 419–36.
5. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem 1997; 66: 61–92.
6. Shereda RD, Kozlov AG, Lohman TM, Cox MM, Keck JL. SSB as an organizer/mobilizer of genome maintenance complexes. Crit Rev Biochem Mol Biol 2008; 43: 289–318.
7. Latchman DS. Transcription factors: an overview. Int J Biochem Cell Biol 1997; 29: 1305–12.
8. Lee T, Young R. Transcription of eukaryotic protein-coding genes. Annu Rev Genet 2000; 34: 77–137.
9. Stegmaier P, Kel AE, Wingender E. Systematic DNA-binding domain classification of transcription factors. Genome Inform 2004; 15: 276–86.
10. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 2006; 34: D108–10.
11. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science 1998; 281: 60–3.
12. Read D. The power of news: the history of Reuters, 2nd ed., Oxford: Oxford University Press, 1999.
13. Matlashewski G, Lamb P, Pim D, Peacock J, Crawford L, Benchimol S. Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. EMBO J 1984; 3: 3257–62.
14. Isobe M, Emanuel BS, Givol D, Oren M, Croce CM. Localization of gene for human p53 tumour antigen to band 17p13. Nature 1986; 320: 84–5.
15. McBride OW, Merry D, Givol D. The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). Proc Natl Acad Sci USA 1986; 83: 130–4.
16. Kern SE, Kinzler KW, Baker SJ, Nigro JM, Rotter V, Levine AJ, Friedman P, Prives C, Vogelstein B. Mutant p53 proteins bind DNA abnormally in vitro. Oncogene 1991; 6: 131–6.
17. Vousden KH, Lane DP. p53 in health and disease. Nat Rev Mol Cell Biol 2007; 8: 275–83.
18. Menendez D, Inga A, Resnick MA. The expanding universe of p53 targets. Nat Rev Cancer 2009; 9: 724–37.
19. El-Deiry WF, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B. Definition of a consensus binding site for p53. Nat Genet 1992; 1: 45–9.

20. El-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, Vogelstein B. WAF1, a potential mediator of p53 tumor suppression. Cell 1993; 75: 817–25.
21. Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell 1993; 75: 805–16.
22. Venot C, Maratrat M, Dureuil C, Conseiller E, Bracco L, Debussche L. The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression. EMBO J 1998; 17: 4668–79.
23. Harms KL, Chen X. The C terminus of p53 family proteins is a cell fate determinant. Mol Cell Biol 2005; 25: 2014–30.
24. Levrero M, De Laurenzi V, Costanzo A, Gong J, Wang JY, Melino G. The p53/p63/p73 family of transcription factors: overlapping and distinct functions. J Cell Sci 2000; 113: 1661–70.
25. Alberts B, Johnson AD, Lewis JA, Raff M, Roberts K, Walter P. DNA replication mechanisms: DNA topoisomerases prevent DNA tangling during replication. In: Molecular biology of the cell, 4th ed., New York: Garland Science, Chapter 5, 2002.
26. Weigel C, Schmidt A, Ruckert B, Lurz R, Messer W. DnaA protein binding to individual DnaA boxes in the Escherichia coli replication origin, oriC. EMBO J 1997; 16: 6574–83.
27. Berg J, Tymoczko J, Stryer L, Clarke N. DNA replication of both strands proceeds rapidly from specific start sites. In: Biochemistry, 5th ed., New York: W.H. Freeman and Company, Chapter 27, Section 4, 2002.
28. Lodish HF. Molecular cell biology, 4th ed., New York: W.H. Freeman, 2000.
29. Stillman BM, De Pamphilis L. Comparison of DNA replication in cells from prokarya and eukarya. New York: Cold Spring Harbor Laboratory Press, 1996.
30. Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem 2002; 71: 333–74.
31. Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 2007; 317: 127–30.
32. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res 2008; 18: 148–61.
33. Barry ER, Bell SD. DNA replication in the archaea. Microbiol Mol Biol Rev 2006; 70: 876–87.
34. Weaver JR, Kugel JF, Goodrich JA. The sequence at specific positions in the early transcribed region sets the rate of transcript synthesis by RNA polymerase II in vitro. J Biol Chem 2005; 280: 39860–9.
35. Alberts B, Johnson AD, Lewis JA, Raff M, Roberts K, Walter P. DNA replication mechanisms. In: Molecular biology of the cell, 4th ed., New York: Garland Science, 2002.
36. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol 2002; 3: 430–40.
37. Miyata T, Oyama T, Mayanagi K, Ishino S, Ishino Y, Morikawa K. The clamp-loading complex for processive DNA replication. Nat Struct Mol Biol 2004; 11: 632–6.
38. Demple B, Herman T, Chen DS. Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes. Proc Natl Acad Sci USA 1991; 88: 11450–4.
39. Liu Y, Prasad R, Beard WA, Kedar PS, Hou EW, Shock DD, Wilson SH. Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase β. J Biol Chem 2007; 282: 13532–41.
40. Braithwaite EK, Kedar PS, Lan L, Polosina YY, Asagoshi K, Poltoratsky VP, Horton JK, Miller H, Teebor GW, Yasui A, Wilson SH. DNA polymerase lambda protects mouse fibroblasts against oxidative DNA damage and is recruited to sites of DNA damage/repair. J Biol Chem 2005; 280: 31641–7.
41. Kao GD, Jiang Z, Fernandes AM, Gupta AK, Maity A. Inhibition of phosphatidylinositol-3-OH kinase/Akt signaling impairs DNA repair in glioblastoma cells following ionizing radiation. J Biol Chem 2007; 282: 21206–12.
42. Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. Chem Rev 2006; 106: 302–23.
43. Larrea AA, Lujan SA, Kunkel TA. SnapShot: DNA mismatch repair. Cell 2010; 141: 730.e1.
44. Griffiths AJF. Introduction to genetic analysis, 9th ed., New York: W.H. Freeman and Co., 2008.
45. de Waard H, de Wit J, Andressoo J-O, van Oostrom CTM, Riis B, Weimann A, Poulsen HE, van Steeg H, Hoeijmakers JHJ, van der Horst GTJ. Different effects of CSA and CSB deficiency on sensitivity to oxidative DNA damage. Mol Cell Biol 2004; 24: 7941–8.
46. D'Amours D, Jackson SP. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. Nat Rev Mol Cell Biol 2002; 3: 317–27.
47. Williams RS, Moncalian G, Williams JS, Yamada Y, Limbo O, Shin DS, Groocock LM, Cahill D, Hitomi C, Guenther G, Moiani D, Carney JP, Russell P, Tainer JA. Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. Cell 2008; 135: 97–109.
48. Mimitou EP, Symington LS. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature 2008; 455: 770–4.
49. Sung JH, Hong SS, Ahn SH, Li H, Seo SY, Park CH, Park BS, Chung SJ. Mechanism for increased bioavailability of tacrine in fasted rats. J Pharm Pharmacol 2006; 58: 643–9.
50. Helleday T, Lo J, van Gent DC, Engelward BP. DNA double-strand break repair: from mechanistic understanding to cancer treatment. DNA Repair 2007; 6: 923–35.
51. Pfeiffer P, Thode S, Hancke J, Vielmetter W. Mechanisms of overlap formation in nonhomologous DNA end joining. Mol Cell Biol 1994; 14: 888–95.
52. Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Mol Cell Biol 1996; 16: 2164–73.
53. Boulton SJ, Jackson SP. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone repair pathways. EMBO J 1996; 15: 5093–103.
54. Wilson TE, Lieber MR. Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase β (Pol4)-dependent pathway. J Biol Chem 1999; 274: 23599–609.
55. Budman J, Chu G. Processing of DNA for nonhomologous end-joining by cell-free extract. EMBO J 2005; 24: 849–60.
56. Espejel S, Franco S, Rodriguez-Perales S, Bouffler SD, Cigudosa JC, Blasco MA. Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres. EMBO J 2002; 21: 2207–19.
57. Guirouilh-Barbat J, Huck S, Bertrand P, Pirzio L, Desmaze C, Sabatier L, Lopez BS. Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol Cell 2004; 14: 611–23.

58. Walker JR, Corpina RA, Goldberg J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature 2001; 412: 607–14.
59. Palmbos PL, Wu D, Daley JM, Wilson TE. Recruitment of Saccharomyces cerevisiae Dnl4-Lif1 complex to a double-strand break requires interactions with Yku80 and the Xrs2 FHA domain. Genetics 2008; 180: 1809–19.
60. Yano K, Morotomi-Yano K, Wang SY, Uematsu N, Lee KJ, Asaithamby A, Weterings E, Chen DJ. Ku recruits XLF to DNA double-strand breaks. EMBO Rep 2008; 9: 91–6.
61. Ma Y, Pannicke U, Schwarz K, Lieber MR. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 2002; 108: 781–94.
62. Nick McElhinny SA, Ramsden DA. Sibling rivalry: competition between Pol X family members in V(D)J recombination and general double strand break repair. Immunol Rev 2004; 200: 156–64.
63. Daley JM, Laan RL, Suresh A, Wilson TE. DNA joint dependence of pol X family polymerase action in nonhomologous end joining. J Biol Chem 2005; 280: 29030–7.
64. Wilson TE, Grawunder U, Lieber MR. Yeast DNA ligase IV mediates non-homologous DNA end joining. Nature 1997; 388: 495–8.
65. Ahnesorg P, Smith P, Jackson SP. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. Cell 2006; 124: 301–13.
66. Buck D, Malivert L, de Chasseval R, Barraud A, Fondaneche MC, Sanal O, Plebani A, Stephan JL, Hufnagel M, le Deist F, Fischer A, Durandy A, de Villartay JP, Revy P. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell 2006; 124: 287–99.
67. Callebaut I, Malivert L, Fischer A, Mornon JP, Revy P, de Villartay JP. Cernunnos interacts with the XRCC4 x DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1. J Biol Chem 2006; 281: 13857–60.
68. Riballo E, Woodbine L, Stiff T, Walker SA, Goodarzi AA, Jeggo PA. XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. Nucleic Acids Res 2009; 37: 482–92.
69. Wang M, Wu W, Rosidi B, Zhang L, Wang H, Iliakis G. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res 2006; 34: 6170–82.
70. Mladenov E, Iliakis G. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. Mutat Res 2011; 711: 61–72.
71. Yan CT, Boboila C, Souza EK, Franco S, Hickernell TR, Murphy M, Gumaste S, Geyer M, Zarrin AA, Manis JP, Rajewsky K, Alt FW. IgH class switching and translocations use a robust non-classical end-joining pathway. Nature 2007; 449: 478–82.
72. Gilfillan S, Dierich A, Lemeur M, Benoist C, Mathis D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. Science 1993; 261: 1175–8.
73. Komori T, Okada A, Stewart V, Alt FW. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. Science 1993; 261: 1171–5.
74. Jung D, Alt FW. Unraveling V(D)J recombination; insights into gene regulation. Cell 2004; 116: 299–311.
75. Boulton SJ, Jackson SP. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. EMBO J 1998; 17: 1819–28.
76. Youngson RM. Collins dictionary of human biology. Glasgow: HarperCollins, 2006.
77. Cox M, Nelson D, Lehninger A. Principles of biochemistry. San Francisco, CA: W.H. Freeman, 2005.
78. Hartl D, Freifelder D, Snyder L. Basic genetics. Boston, MA: Jones and Bartlett Publishers, 1988.
79. Bhasin M, Reinherz EL, Reche PA. Recognition and classification of histones using support vector machine. J Comput Biol 2006; 13: 102–12.
80. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 1997; 389: 251–60.
81. Farkas D. DNA simplified: the hitchhiker's guide to DNA. Washington, DC: AACC Press, 1996.
82. Lennox RW, Oshima RG, Cohen LH. The H1 histones and their interphase phosphorylated states in differentiated and undifferentiated cell lines derived from murine teratocarcinomas. J Biol Chem 1982; 257: 5183–9.
83. Talasz H, Helliger W, Puschendorf B, Lindner H. In vivo phosphorylation of histone H1 variants during the cell cycle. Biochemistry 1996; 35: 1761–7.
84. Sarg B, Helliger W, Talasz H, Förg B, Lindner HH. Histone H1 phosphorylation occurs site-specifically during interphase and mitosis. J Biol Chem 2006; 281: 6573–80.
85. Bustin M. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. Mol Cell Biol 1999; 19: 5237–46.
86. Laudet V, Stehelin D, Clevers H. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res 1993; 21: 2493–501.
87. Boonyaratanakornkit V, Melvin V, Prendergast P, Altmann M, Ronfani L, Bianchi ME, Taraseviciene L, Nordeen SK, Allegretto EA, Edwards DP. High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. Mol Cell Biol 1998; 18: 4471–87.
88. Pierantoni GM, Rinaldo C, Mottolese M, Di Benedetto A, Esposito F, Soddu S, Fusco A. High-mobility group A1 inhibits p53 by cytoplasmic relocalization of its proapoptotic activator HIPK2. J Clin Invest 2007; 117: 693–702.
89. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. EMBO J 1997; 16: 2665–70.
90. Zappavigna V, Falciola L, Helmer-Citterich M, Mavilio F, Bianchi ME. HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. EMBO J 1996; 15: 4981–91.
91. Lotze MT, Tracey KJ. High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal. Nat Rev Immunol 2005; 5: 331–42.
92. Rajeswari M, Jain A. High-mobility-group chromosomal proteins, HMGA1 as potential tumour markers. Curr Sci 2002; 82: 838–44.
93. Lu L, Wang Y. Immunoprecipitation alert: DNA binding proteins directly bind to protein A/G without any antibody as the bridge. Cell Cycle 2008; 7: 417–8.
94. Cho Y, Goprina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 1994; 265: 346–55.

Received February 28, 2011; accepted May 23, 2011